Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v9n4p760

Critical comparison of the main methods for the technical efficiency

By Scippacercola, Sepe

Published: 15 December 2016

Electronic Journal of Applied Statistical Analysis
Vol. 09, Issue 04, December 2016, 760-780
DOI: 10.1285/i20705948v9n4p760

Critical comparison of the main methods for the technical efficiency

Sergio Scippacercola*and Enrica Sepe

Università degli studi di Napoli "Federico II",
Dipartimento di Economia, Management, Istituzioni
Via Cinthia, Napoli (Italy)

Published: 15 December 2016

The Technical Efficiency is a basic tool to determine the factors that slow down the production. TE aims at evaluating and comparing the operating performance of a set of production units, such as Companies, Offices, Hospitals, Banks, Schools, Transport Systems, etc. This paper, after an overview of the literature regarding the methodologies for measuring the Technical Efficiency, compares critically the two main approaches, the Data Envelopment Analysis (DEA) and the Stochastic Frontier Analysis (SFA). These methodologies are also discussed within an original application that targets to study the efficiency of European Countries with respect to the Gross Domestic Product (GDP).

keywords: Technical Efficiency, Stochastic Frontier Analysis, Data Envelopment Analysis, Gross Domestic Product.

1 Introduction

There are various methods for evaluating and comparing the operating performance of a set of production units, such as Companies, Departments, Hospitals, Bank Branches, Transport Systems, etc. The performance of a production unit may be measured with respect to multiple dimensions. All methodologies that tend to evaluate the productive units are based on productivity indicators otherwise known as Technical Efficiency (TE), which provides a measurement characterizing the activity of the units to be compared. This measure is defined according to the results produced by each unit (output) and the

*Corresponding author: sergio.scippacercola@unina.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

resources used to achieve those results (inputs or factors of production). For example, if the units are bank branches, the outputs can be the number of current accounts, the number of checks changed, the number of mortgages taken out in the last year, etc. while the input may be the number of employees, the area available, the number of weekly hours of opening, etc.

This paper reviews the main methodologies for measuring TE. Section 2 introduces the Technical Efficiency while Section 3 presents the methods for measuring the TE. Section 4 presents more in detail the Data Envelopment Analysis (DEA), while Section 5 details the Stochastic Frontier Analysis (SFA). Section 6 discusses two approaches within an original study that targets to measure the efficiency of some European Countries, with respect to the Gross Domestic Product (GDP). Section 7 highlights the differences among DEA, SFA and OLS regression, which are critically evaluated in Sections 8 and Section 9, with a discussion on their pros and cons. Finally, Section 10 concludes the paper with final considerations.

2 Technical Efficiency

Technical Efficiency (TE) is a basic tool for determining the factors that slow down production. The literature on the measurement of Technical Efficiency provides a range of methodologies.

In this section we introduce the main concepts on the assessment of technical efficiency to compare many productive units homogeneous amongst themselves. The TE is an indicator that provides a measurement on the basis of the results produced (outputs) from each production unit to the resources used (inputs). The production unit will be called Decision Making Unit (DMU) following the main literature in this domain. To evaluate the efficiency of n DMU, where each unit produces only one output employing a single input, let y be the value of output produced and x the value of input employed, the TE of the generic DMU is defined by: $TE = y/x$.

A *production function f or efficiency frontier* is defined as the schedule of the maximum amount of output that can be produced from a specified set of input, given the existing technology (Battese, 1992): f represents the input/unit output combinations possible when the available technology is efficiently utilized, i.e. the unit is an *isoquant* of the efficient producer (McMillan and Chan, 2006).

For example, considering as input the number of employees x and the sales figures as output, with reference to Figure 1, y_A and y_B represent the outputs produced by the DMUs A and B, the slope of the each straight line joining a point with the origin of the axes is the value of efficiency associated with each DMU. The straight line with the maximum slope represents the efficient frontier f. Each point on f as B is considered efficient. The TE of A can be measured as y_A /x, while TE is the percentage of the output that the DMU A could produce if it were fully efficient. Alternatively, the *Technical Inefficiency (TI)* of A $TI = (y_B$ -y_A $)/y_B$ $(0 \leq TE \leq 1)$ is also defined and represents the percentage to become efficient.

Technical Efficiency can be measured by two main approaches, namely *input oriented* and *output oriented*. The TE input oriented is the ratio x^*/x_i where x^* is the ideal amount of input that the unit should employ if it were used efficiently and x_i is the amount employed by *i-th* DMU (Figure 2). The TE output oriented is the ratio y_i/y^*, where y^* is the ideal amount of output that the unit should produce if it were used efficiently and y_i is the quantity produced by *i-th* DMU (Figure 3). The efficiency frontier provides guidelines for the improvement of inefficient units: it identifies the level of output or input achieved by the units in terms of efficiency. The frontiers are different in the case of single input single output (Figure 1), orientation to the inputs (Figure 2) and orientation output (Figure 3).

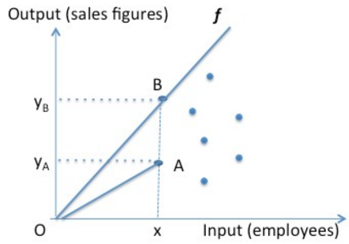

Figure 1: f efficient frontier

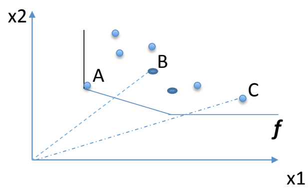

Figure 2: TE input oriented

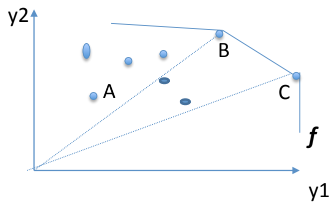

Figure 3: TE output oriented

3 The methods for evaluating technical efficiency

The methods for measuring TE can be classified as parametric (deterministic or stochastic) and non-parametric (deterministic or stochastic)(Figure 4):

3.1 Parametric: Frontier Deterministic Frontier, Ordinary least-squares models and Stochastic Frontiers

Each Econometric estimation of parametric functions has a precise mathematical form which is not very simple and straightforward to identify.

The **Deterministic Frontiers** (Aigner and Chu, 1968) are parametric functions in which the deviation of an observation from its theoretical maximum is attributed exclusively to the inefficiency of the firm and does not take into account casual shocks.

The **Ordinary Least-Squares** (OLS) regression is the most well known generalized linear modelling technique that may be used to model a single response variable, which has been recorded on at least one interval scale. The technique may be applied to single or multiple explanatory variables and also categorical explanatory variables that have been appropriately coded (Moutinho and Hutcheson, 2011). The method of least squares (OLS) produces a line that minimizes the sum of the squared vertical distances from the line to the observed data points (DMUs).

The **Stochastic Frontier** (SF) (Aigner et al., 1977; Battese and Corra, 1977; Meeusen and Van den Broeck, 1977; Van den Broeck et al., 1994) assumes that it is not possible to fully specify the function and allows for random noise. The Stochastic Frontier, therefore, takes into account the random component.

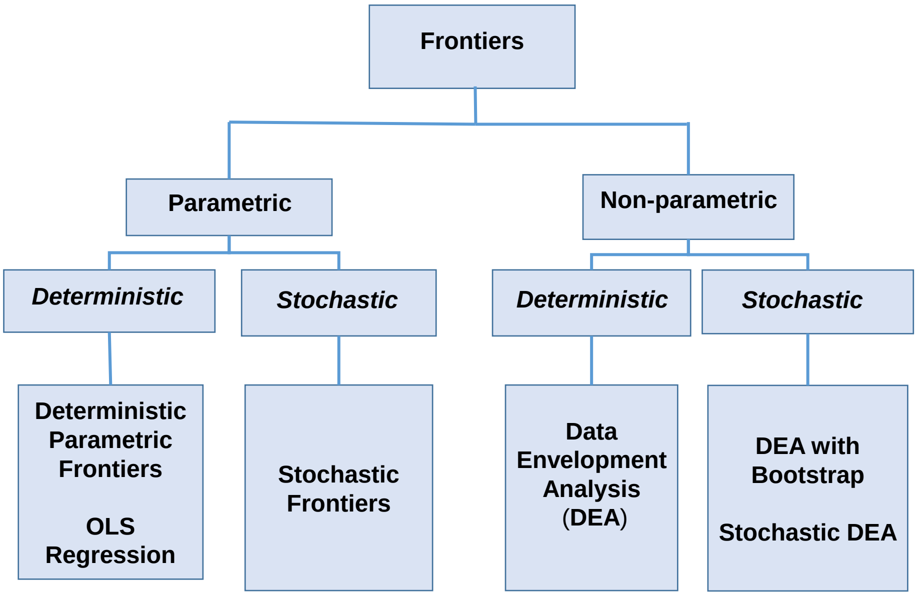

Figure 4: Methods for evaluating Technical Efficiency

3.2 Programming or Non-Parametric models: DEA, Stochastic DEA and Bayesian model

In the non-parametric frontier one excludes the participation in the efficiency of random components and only a few non-affirmed methods (DEA with bootstrap and Stochastic DEA) take into account some random components. The most famous is **DEA** (Charnes et al., 1978; Charnes et al., 1981) that is a nonparametric deterministic model, while **DEA with Bootstrap** (Simar and Wilson, 1998) and **Stochastic DEA** (Sengupta, 1987) are non-parametric stochastic model. To use Stochastic DEA it is necessary to provide information about the expected values and variances of all variables, as well as probability levels at which feasibility constraints are satisfied (Porcelli, 2009).

Finally, the **Bayesian model** treats the uncertainty regarding which sampling model to use by mixing up a number of competing inefficiency distributions proposed in literature with a model probabilities as weights. The choice of a particular distribution for the inefficiency term most favored by the data can be made using Bayes factors or posterior odds ratio as a criterion for model selection. The Bayesian approach is a response to overcome the criticism of imposing a priori sampling distributions on the efficiency-related random variable in the SFA (Orbanz and Teh, 2010).

The main Approaches to technical efficiency from 1951 are in Table 1. Most success was achieved only by the DEA and SFA that are treated in more detail in the following paragraphs and then are compared to their advantages and disadvantages.

Table 1: Brief list of methods for the technical efficiency

Year	Author	Method
1957	FARREL	Efficiency index
1968	AIGNER, CHU	Frontier Deterministic Frontiers
1977	AIGNER, LOVELL, SCHMIDT	SFA
1977	BATTESE, CORRA	SFA
1977	MEEUSSEN, VAN DEN BROECK	SFA
1978	CHARNES, COOPER, RHODES	DEA-CCR (constant returns to scale)
1984	BANKER, CHARNES, COOPER	DEA-BCC(variable returns to scale)
1987	SENGUPTA	SDEA
1988	BATTESE, COELLI	Generalized frontier production
1993	FRIED, LOVELL	Productive Efficiency
1994	VAN DEN BROECK et al.	Bayesian approach
1995	OLESEN	DEA
1998	SIMAR, WILSON	DEA with Bootstrap
1998	COELLI et al.	DEA
2005	COELLI, et al.	DEA
2010	ORBANZ, TECH	Bayesian Approach

4 Data Envelopment Analysis

4.1 Brief History

In 1978, Charnes, Cooper and Rhodes (CCR) (Charnes et al., 1978), starting from an efficiency index of Farrell (1957), introduced Data Envelopment Analysis as a "mathematical programming model applied to observational data". DEA consists in a new algorithmic method to the efficiency measurement of the Data Making Units (DMU) for constant returns to scale (DEA CRS), where all DMUs are operating at their optimal scale. DEA allows multiple inputs and outputs to be considered at the same time without any assumption made on data distribution.

Banker, Charnes and Cooper (Banker et al., 1984) introduced the variable returns to the scale efficiency measurement model, allowing the breakdown of efficiency into technical and scale efficiencies in DEA.

4.2 The Model

If l units produce multiple output using several factors of input, Technical Efficiency $TE_j(j = 1, 2, ..., l)$ is measured in terms of a proportional change among inputs and outputs:

$$TE_j(v, w) = \frac{\sum_{k=1}^{m} w_k y_{kj}}{\sum_{i=1}^{n} v_i x_{ij}} \tag{1}$$

where:

- x_{ij} $(i = 1, 2, ..., n; j = 1, 2, ..., l)$ is the input i.e. the amount of the input used by the DMU_j;
- v_i is the weight associated with the input;
- y_{kj} are the outputs $(k = 1, 2, ..., m)$ of the DMU_j; w_k are the weights associated with the outputs.

DEA assesses the efficiency of each unit by the weighting system that is most appropriate to the DMU. The objective function is to maximize the TE for DMU_j according to the weights v and w:

$$max \quad TE_j(v, w) \quad (j = 1, 2, ..., l) \tag{2}$$

The maximization is subject to the following constraints:

- no DMU can operate beyond the production possibility set (3) i.e. that the efficiency value for each unit is not more than one:

$$TE_j(v, w) \leq 1 \quad (j = 1, 2, ..., l) \tag{3}$$

- the weights are non negative:

$$v_i, w_k \geq 0 \quad (i = 1, 2, ..., n; k = 1, 2, ...m) \tag{4}$$

The model (2) becomes linear implying that the weighted sum of the inputs is equal to 1, in which case the model is called CCR input-oriented:

$$max\, TE_j = \sum_{k=1}^{m} w_k \quad y_{kj} \tag{5}$$

$$\sum_{k=1}^{m} w_k y_{kj} - \sum_{i=1}^{n} v_i x_{ij} \leq 0 \quad (j = 1, 2, ..., l) \tag{6}$$

$$v_i,\ w_k \geq 0 \quad (i = 1, 2, ..., n; k = 1, 2, ...m) \tag{7}$$

For the model CCR input-oriented it is possible a dual formulation minimizing (1). Denoting with TE^* the optimum value of the objective function at the optimal solution (v^*, w^*), the DMU is said efficient if $TE^* = 1$ and if there exists at least one optimal solution (v^*, w^*), where $v^* > 0$ and $w^* > 0$. The model assumes that the comparison operating unit returns to scale are constant[1].

DEA model aims to choose the system of weights for the input and the output through a mathematical programming model: an input-oriented model, which minimizes the input while satisfying at least the given output level; and an output-oriented model, which maximizes the weighted sum of the outputs. In particular for each DMU the input-oriented efficiency is the relationship between the ideal amount x^* and the x_j quantity actually applied. Similarly, the output-oriented efficiency is the ratio between the y_j quantity output and the ideal amount y^* that it should produce in conditions of efficiency.

Making efficient an inefficient unit means identifying the resources with which to bring the efficiency units near the border of efficiency. The presence of *slacks* indicates that the DMU is not efficient and would therefore be possible to maintain the same level of production by reducing the resources used. If $TE = 1$ is difficult to determine at what extent the value of efficiency is due to the high level of efficiency or to the selection of the optimal structure of weights. For further details and a different version of the method, see Cooper et al., 1996, Thanassoulis, 2001 and Ray, 2004.

5 Stochastic Frontier Analysis

5.1 Brief History

The SFA has its roots with the publication *Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error* by Meeusen and van den Broeck (Meeusen and Van den Broeck, 1977) and then the other two still in 1977, namely Aigner, Lovell, Schmidt (Aigner et al., 1977) and Battese and Corra (Battese and Corra, 1977). It is interesting to start with the original statement of Aigner, Lovell and Schimdt, according to which: "The theoretical definition of a production function expressing the maximum amount of output obtainable from given input bundled with fixed technology has been accepted for many decades. And for almost as long, econometricians have been estimating average production functions. It has only been since the pioneering work of Farrell (1957) that serious consideration has been given to the possibility of estimating so-called frontier production functions, in an effort to bridge the gap between theory and empirical work. For a variety of reasons these efforts have not been completely successful. In this paper we suggest a new approach to the estimation of frontier production functions. This involves the specification of the error term as being made up of two components, *one normal and the other from a one-sided distribution.* This approach enables us to

[1]The yields of scale expressing the variation in the amount of output occurred due to changes in the amount employed in the input. If the returns to scale are constant, with an increase in input corresponds to an increase ino output in the same proportion. If the returns to scale are variable: an increase in inputs does not result in a proportional change in the outputs.

overcome some of the major shortcomings of previous work in the area" (Aigner et al., 1977).

The model has been over time a topic for discussion regarding the sensitivity of the results obtained with respect to the type of distribution specified for the component of technical inefficiency. Developments for the inefficiency error component are in Green (Greene, 1980) (gamma distribution); Stevenson (1980) (gamma and truncated normal distributions); Lee (1983) (four-parameter- Pearson family of distributions); and in addition Førsund et al. (1980), Bauer (1990). In the following, many other forms of distribution have been proposed for the component of technical inefficiency but most of the research is almost always aimed at the half normal or exponential distributions.

5.2 The Model

The SFA model depends on specifying a functional form f which relates the outputs to the inputs. It is then necessary to estimate the parameters of f function subject to certain assumptions about the distribution of the residuals. The usual model of the Stochastic Frontier Analysis is (Coelli et al., 2005):

$$\ln y_j = \mathbf{x}_j'\boldsymbol{\beta} + v_j - u_j \quad (j = 1, 2, ..., l). \tag{8}$$

where:

- y_j is the scalar output of the $i-th$ DMU;
- $\mathbf{x_j}$ is a vector of inputs;
- $\boldsymbol{\beta}$ is a vector of $k+1$ technology parameters to be estimated;
- v_j represents the first error component, i.e. all events which are not under control (such as random noise, omitted variables, etc.); $v_i \approx iid\ N(0, \sigma_v^2\ )$ is the noise or error term or the measure of effects independent of the DMU; v_i is homoskedastic;
- u_j is the second error component or inefficiency (i.e., all events which are under control); u_j is a non-negative random variable measuring the technical inefficiency with half-normal either normal-truncated model (Stevenson, 1980) or exponential or gamma (Greene, 1980);
- v_j and u_j are distributed independently of each other and of the regressors.

Many other forms of distribution have been proposed for the component of technical inefficiency but most of the researchers is addressed almost always distributions half normal or exponential. Kumbhakar and Lovell (Kumbhakar, 2000) pointed out that the Stochastic Frontier Model yields a clear "separation of shock due to variable inputs from the effects of environmental variables on the production (and thus also on the efficiency)".

Model (8) has been over time a topic for discussion regarding the sensitivity of the results obtained with respect to the type of distribution specified for the component of technical inefficiency. Model (8), without one of the error components, generates other families of models: a deterministic production frontier ($if\ v_i = 0$) and stochastic production function model ($if\ u_i = 0$). The deterministic production frontier can be estimated using COLS (Winsten, 1957), that does not require any assumptions about the functional form of u_i or MOLS (Richmond, 1974) and MLE (Afriat, 1972).

The *Deterministic Production Frontier* approach (Schmidt et al., 1976) is:

$$y_i = f(\mathbf{x_j}; \boldsymbol{\beta}) - u \quad u \geq 0 \tag{9}$$

where $\boldsymbol{\beta}$ is a vector of parameters estimated by one of the following methods: Linear (Aigner and Chu, 1968), COLS (Olson et al., 1980), ML (Schmidt et al., 1976) and u ($u \geq 0$) represents the residual or inefficiency. In this case the Technical Efficiency is:

$$TE_j = \frac{y_j}{f(\mathbf{x_j}; \boldsymbol{\beta})} \quad 1 \geq TE_j > 0 \tag{10}$$

The estimates of the parameters have no statistical properties (Schmidt, 1985) and the residuals are taken as measures of efficiency. The two most used and usual forms of f in (8) are the Cobb-Douglas (Cobb and Douglas, 1928) and Translog, and among other versions of SFA we find Diewert (Diewert, 1971), Christensen et al. (1973); Gong and Sickles (1992), Gallant (Fourier flexible form) (Gallant, 1981).

This model depends on specifying a functional form f, which relates the outputs to the inputs. Then it is necessary to estimate the parameters of f function subject to certain assumptions about the distribution of the residuals.

To estimate whether there is a stochastic frontier we can use a gamma index (Battese and Corra, 1977):

$$\gamma \ = \ \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2} \quad 0 \leq \gamma \leq 1 \tag{11}$$

The γ parameter can have values between zero and one; in the case where it is equal to zero, it means that the contribution to the total variability of the inefficiency is null and this implies that in the system there is not true inefficiency but only stochastic errors: the parameters of (5) can be estimated using OLS. Conversely, in the case where γ is equal to 1, it means that all the variability is due to the inefficiency that is the businesses are not affected by exogenous shocks (the model is deterministic with no noise). If γ *is close to* 1 it indicates that the deviations from the frontier are due mostly to technical inefficiency.

The parameters of stochastic frontier function are estimated by the maximum likelihood method. The Technical Efficiency (TE) of *j-th* DMU is the ratio of realised output to the stochastic frontier output:

$$ln\, TE_j = ln\, y_j - ln\, y_j^* = ln \frac{y_j}{y_j^*} - u_j \quad (j = 1, 2, ..., l) \quad (0 \leq TE_j \leq 1). \tag{12}$$

For further details of the method, see Aigner and Chu (1968), Battese and Corra (1977).

6 An application of SFA and DEA

6.1 The Data

To provide an application of the SFA and DEA on a real case study, we present a study on the TE of thirty European Countries to produce Gross Domestic Product (GDP), in the year 2013. We also analyze which variables, among the examined, can be responsible for the inefficiency. The data have been provided by Eurostat and are the basis both for SFA and DEA. The data covers attributes regarding economic resources.

We consider as **output variable** the GDP (y_{i1}) that is an indicator for a $nation's$ economic situation, where i refers to the *i-th* Country. It reflects the total value of all goods and services produced, less the value of goods and services used for intermediate consumption in their production. Expressing GDP in PPS (purchasing power standards) eliminates differences in price levels between countries, and calculations on a per head basis allows for the comparison of economies significantly different in absolute size (Eurostat).

There is a large set of input variables that can potentially explain the differences of technical efficiency among the Countries. After significance tests, the following **variables of input** have been kept on the list of the potential determinants of TE, that represent some characteristics of the Country with respect to the GDP (i refers to the *i-th* Country):

- x_{i1} is an indicator that represents the value of export of goods and services divided by the imports of goods and services. Values higher than one indicate a positive trade balance whereas values smaller than one indicate a negative trade balance.
- x_{i2} is the value of exports of goods and services divided by the GDP in current prices.
- x_{i3} is the value of imports of goods and services divided by the GDP in current prices.
- x_{i4} is defined as the total remuneration (compensation of employees), in cash or in kind, payable by an employer to an employee in return for work done by the latter. In particular, it also includes social contributions paid by the employer.
- x_{i5} are the taxes and subsidies on products are current unrequited payments to or from general government or the Institutions of the European Union that are payable per unit of some good or service produced or transacted. The tax or subsidy may be a specific amount of money per unit of quantity of a good or service, or it may be calculated ad valorem as a specified percentage of the price per unit or value of the goods and services produced or transacted.

6.2 SFA results

We started from the model including all variables and interactions. The choice of the functional form has been taken under the hypothesis of a parsimonious model. The null hypothesis of absence of random technical inefficiency is rejected and thus the Stochastic Frontier Analysis seems appropriate for the data. After verifying the hypothesis of asymmetry present in the residuals of the OLS and after trying several models with different dependent variables, the model of SFA is:

$$\ln(y_{1i}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 \ln(x_{i4}) + \beta_5 x_{i5} + v_i - u_i \tag{13}$$

where i refers to the *i-th* Country. Variables v_i and u_i are defined as described in Section 5.2. We analyzed three models (half-normal, truncated normal and exponential). Half-normal distribution for the efficiency term proved more significant than truncated normal and exponential tested models. Thus, by using the log likelihood values, we have chosen the half-normal model. Table 2 summarizes the main results of the model (13).

The test for statistical significance of the deterministic inefficiency portion of the total error involves the computation of γ ($H_0 : \gamma = 0; H_1 : \gamma \neq 0$). Using the most basic production function form along with a half-normal deterministic inefficiency error, γ is equal to 0.96. The likelihood ratio test statistic for σ_u based on a mixed χ^2 distribution is 2.53 (Table 2). This supports rejection of the null hypothesis: the inefficiency is a significant portion of the total error and SFA is appropriate for the analysis. The model is significant ($Prob > \chi^2 = 0.0000$).

The coefficients of the variables are positive except in the case of x_{i3} (imports of goods and services). The results (Table 2) of the model (13) show that only the input variable compensation of employees (x_{i4}) has a significant impact on the determination of the production frontier of GDP. The input variables that are obstacles to efficiency are the variables of input as imports of goods and services, while the variables export of goods and services (x_{i1} and x_{i2}) improve the efficiency. The maximum value for technical efficiency (TE=1) is not achieved by any Country; however, using the values of TE, we can identify decreasing ranks of efficiency (Table 3).

Table 2: Estimation results of Frontier Production Functions with dependent variable being the GDP

Input Variables	Coefficient	Standard Error	z	$P > \|z\|$	95% Confidence interval	
Constant	6.024604	2.087161	2.89	0.004	1.93384	10.11530
x_{i1}	1.817249	1.243510	1.46	0.144	-.619985	4.25448
x_{i2}	0.029547	.0195756	1.51	0.131	-.008820	.06791
x_{i3}	-.0292432	.0234128	-1.25	0.212	-.075131	.01664
x_{i4}	.1984719	.0608725	3.26	0.001	.079164	.31777
x_{i5}	.0177708	.0232014	0.77	0.444	-.027703	.06324
σ_u	.6783084	.13512131				
σ_v	.1390061	.0896505				
γ	0.96					
Log likelihood	-15.168287	$Prob > \chi^2 = 0.0000$				
Likelihood-ratio test of $\sigma_u = 0$	$\overline{\chi}^2(01) = 2.53$	$Prob \geq \overline{\chi}^2 = 0.0000$				

6.3 DEA results

In its simplest form, it is constructed from the set of relevant inputs and desirable outputs of the process, together with some basic, standard assumptions on the nature of the production possibilities. Thus, by analyzing the input/output data of set of similar units (e.g. countries, stores), DEA identifies:

- the efficient frontier consisting of the best practice units;
- the efficiency measures for each DMU that reflect their distance from the frontier (this measure is equal to 1 for efficient DMUs, and less than one otherwise, as illustrated in Section 4.1 and 4.2);
- an efficient reference set (a small subset of efficient units "closest" to the unit under evaluation) for each inefficient DMU (Table 4) (Schaffnit et al., 1997).

In this study we used an input-oriented DEA model. This is a natural choice since the branches have, in general, no direct control over the amount of services that their customers require. These models yield to scores and targets consistent with the *i-th* Country's objective of improving its efficiency at the current levels of service. As will be explained later (Section 8 and 9), the choice of radial models also allows us to work within a global frame-work for model comparison. To investigate scale efficiency, we use models with constant returns to scale (CSR).

Table 3: Technical efficiency, by SFA, and rank of European countries with output variable being the GDP

Country	Technical Efficiency	Rank	Country	Technical Efficiency	Rank
Finland	.9398	1	Netherlands	.6011	16
Cyprus	.8960	2	Lithuania	.5595	17
Belgium	.8782	3	Slovenia	.5566	18
Denmark	.8627	4	Portugal	.5230	19
Luxembourg	.8443	5	Ireland	.4985	20
Switzerland	.8414	6	Italy	.4774	21
Austria	.8394	7	Slovakia	.4742	22
Norway	.8306	8	Germany	.4634	23
Malta	.8026	9	Spain	.4415	24
United Kingdom	.7594	10	Croatia	.4090	25
France	.7559	11	Czech Republic	.4037	26
Greece	.7526	12	Romania	.3003	27
Sweden	.7436	13	Bulgaria	.2890	28
Latvia	.7043	14	Poland	.2803	29
Estonia	.6992	15	Hungary	.2608	30

Table 4 gives a summary of the results for the basic CCR model. The scores tell us that this model identifies a DMU efficient for Finland ($TE^* = 1$ as indicated in section 4), Cyprus, Belgium, Luxembourg, Switzerland, Norway, Malta, United Kingdom, France, Greece, Italy and Spain. Even for Countries with value of Theta = 1, DEA gives *slacks*: however, these were not reported in the results for reasons of space. Table 4 also gives the ranking for each country classified on the basis of presence of one or more slacks. The decreasing values of Theta indicate an increasingly inefficiency of the Country. For each country, DEA highlights, with the presence of slacks different from zero, the input variable that should be object of correction by the DMU (i.e. the country) to reach the maximum efficiency level. The efficiencies obtained with the DEA model (Table 4) are almost similar to the results with the SFA model (Table 3).

Table 4: Technical efficiency (Theta), by DEA, and rank of European countries with output variable being the GDP

Country (DMU)	Theta	Rank	Country (DMU)	Theta	Rank
Finland	1	1	Netherlands	.9383	23
Cyprus	1	2	Lithuania	.9678	15
Belgium	1	10	Slovenia	.9226	25
Denmark	.9494	17	Portugal	.9341	24
Luxembourg	1	12	Ireland	.9419	20
Switzerland	1	3	Italy	1	8
Austria	.9444	19	Slovakia	.9408	21
Norway	1	4	Germany	.9393	22
Malta	1	11	Spain	1	9
United Kingdom	1	5	Croatia	.9472	18
France	1	6	Czech Republic	.9086	26
Greece	1	7	Romania	.8910	27
Sweden	.9531	16	Bulgaria	.8794	28
Latvia	.9990	13	Poland	.8590	29
Estonia	.9832	14	Hungary	.8481	30

7 Differences among DEA, SFA and OLS regression

By means of SFA we obtain a continuous stochastic frontier function, unlike DEA that, due to its optimizations, identifies a frontier composed of multiple segments. In Figure 5, by way of example[2], the regression line, the stochastic frontier (continuous line) and the frontier produced by DEA (piece-wise) are represented. The measure of efficiency is normally one of either: the distance between observed and maximum possible output for given inputs (output efficiency); the distance between observed and minimum possible input for given outputs (input efficiency).

If the three lines are different, then the values of technical efficiency can differ among them: some values can be overestimated and others underestimated. In our application Greece, Italy, Spain and Switzerland are overestimated by DEA (Table 4), whereas Denmark and Austria are underestimated by SFA (Table 3).

The regression line obtained according to the available observations (Figure 5) is an alternative to the efficiency frontier. Those are considered excellent units, which are

[2](Re-elaboration of a graph by Cordeiro et al., 2012)

situated above the regression line and you can evaluate the degree of efficiency of each unit depending on the distance from the regression line. The regression line reflects the **average behavior** of the units to be compared, the frontiers drawn from DEA or by SFA identify, however, the **behavior more accurately** and measure the inefficiency of a unit according to the distance that separates it from the border itself.

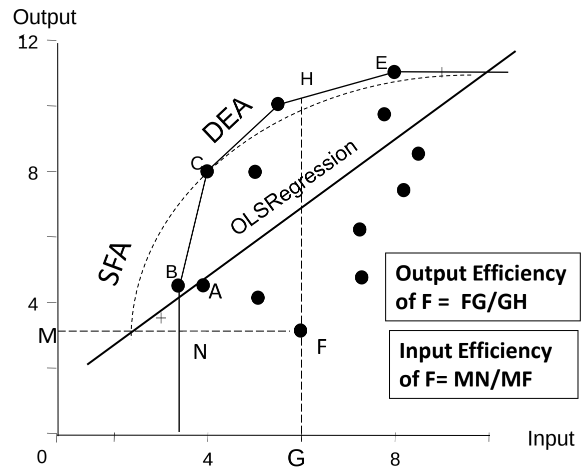

Figure 5: Example of frontiers through DEA and SFA versus OLS Regression

8 Pros and cons of DEA

The main **advantages** in the use of DEA are listed below.

- The presence of *slacks* indicates that the DMU is not efficient and it would therefore be possible to maintain the same level of production by reducing the resources used.
- DEA does not have restrictions on the functional form of the production relationships between I/O. DEA does not require any distributional assumption. DEA uses multiple Input and Output simultaneously (Kalirajan and Shand, 1999). These are some reasons for that DEA is more appealing to the users respect to the SFA.
- Assumes convexity of the production possibility space (Olesen, 1995).
- DEA extract information on the values of the inputs and outputs (*slacks*) to achieve efficiency.
- DEA identifies for each unit an efficient and excellent peer group that consists of the units that are efficient when measured using the optimal weight system for the unit inefficient. The peer groups are a reference to which the inefficient units can inspire you to improve their performance.
- *With the growth and the easiness of the software to develop the analysis (SAS, STATA, etc.), DEA is today widely used as a managerial tool for measuring the performance of public and private organizations.*
- *DEA is preferred, also, when the parametric methods (SFA, OLS) are not applicable for the non-validity of assumptions about the parametric model.*

The main **disadvantages** of DEA are shown below.

- DEA incorporates noise as part of the efficiency score. Using DEA model, efficiency scores are contaminated by omitted variables, measurement errors, and other sources of statistical noise.
- The weights become critical for the evaluation of the efficiency, if $TE_j = 1$ it is difficult to determine at what extent the value of efficiency is due to the high level of efficiency or to selection of the optimal structure of weights.
- DEA is extremely sensitive to the selection of variables and to data errors. In small samples, the DEA efficiency measures are sensitive to the difference between the number of firms: many firms may be seen to be efficient, even though they are not (Seiford, 1996). Indeed, considering the thirty countries examined, nine countries resulted to be efficient ($\theta = 1$), but their efficiency was not confirmed by the SFA.
- An estimate of the efficient frontier yields insufficient information to establish Frontier because it is derived using only marginal data. TE measures are susceptible to extreme observations and measurement errors (Førsund et al., 1980).
- DEA is sensitive to outliers.

9 Pros and cons of the SFA

The main **advantages** in the use of SFA are the following.

- Based on the Theory of the Regression. SFA uses maximum likelihood econometric estimation, the hypotheses can be statistically tested, and SFA is in conformity with production theory.
- SFA offers flexibility in modelling various specific aspects of production (e.g., the risk), the distribution of the random noise term and the inefficiency term.
- By means of SFA no correlation between inefficiency and exogenous variables - Inefficiency only in endogenous variables (Olesen, 1995).
- SFA uses a hypothesized function to calculate estimates of the efficiencies of individual DMUs, SFA can separate random noise from efficiency. SFA uses a hypothesized function to calculate estimates of the efficiencies of individual DMUs.
- Only SFA can separate random noise from efficiency.
- SFA and OLS regression methods reveal overall sample-based information;
- The ability to obtain *specific estimates* for producing efficiency has greatly improved the attraction to the SFA.
- SFA is non-sensitive to outliers.

The main **disadvantages** of SFA are listed below.

- SFA requires a certain specific distributional assumption on firm-specific TE related variables u (Kalirajan and Shand, 1999).
- Do not use simple mathematical forms.
- SFA uses multiple Input and single Output and no assumptions about the form of Technology.

10 Final considerations

"It is asked whether it is reasonable to compare a deterministic method with a stochastic method?" (Førsund, 1992). What is the method to be chosen? We can distinguish **three distinct groups**.

The first group is formed by researchers that measure the technical efficiency with a double measurement if the assumptions of SFA are verified. Then the results are compared to see if the estimates obtained are equal or different.

Whenever the methods do not give similar estimates, the decision maker needs to be able to tell which of the methods is giving closer estimates to the true values.

Another group of studies measures the technical efficiency by adopting SFA, and if the assumptions are verified, they use the DEA to observe if the estimates obtained are identical or different.

The third group, which is most of the people, instead, choose the easiest method to implement which often leads to choosing DEA rather than SFA methods: the results of DEA can be easier to analyse, the performance of the methods is highly dependent upon the data set, which is being analysed.

Using SFA the determinants of efficiency are directly obtained by estimating the production function. In SFA you can use various models, changing the response variable every time, to identify the model that has greater relevance in terms of acceptance. In SFA the hypothesis can be tested very strictly and this is the reason why it is preferable than DEA, specially by statisticians.

Both DEA and SFA analyses are popular methods for assessing relative efficiency. Unfortunately, there is no definitive formula for deciding which to adopt. The decision is a call for judgment. Obviously, a case can be made for each and analysts have chosen to use both (though rarely together). Since each is a viable option, it is logical to check the sensitivity of efficiency outcomes to the method of analysis.

Because the approaches are different, some difference in their outcomes is to be expected. However, because they are alternative approaches to a common problem, their outcomes should be compared as we do here. The methods do not give similar estimates, so the decision maker needs to be able to tell which of the methods is giving closer estimates to the true values. If the estimates are not equal we must know that DEA overestimation is compared to values calculated with SFA.

Finally, we can only affirm that the most frequently used methods for efficiency estimation are DEA in the non-parametric literature and SFA in the parametric literature, while this work sets out only to compare the assumptions that form the basis of the two methods in order to provide critical points of reflection to those who want to measure efficiency.

References

Afriat, S. N. (1972). Efficiency estimation of production functions. *International Economic Review*, pages 568–598.

Aigner, D., Lovell, C. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *journal of Econometrics*, 6(1):21–37.

Aigner, D. J. and Chu, S.-f. (1968). On estimating the industry production function. *The American Economic Review*, 58(4):826–839.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, 30(9):1078–1092.

Battese, G. E. (1992). Frontier production functions and technical efficiency: a survey of empirical applications in agricultural economics. *Agricultural economics*, 7(3):185–208.

Battese, G. E. and Corra, G. S. (1977). Estimation of a production frontier model: with application to the pastoral zone of eastern australia. *Australian journal of agricultural economics*, 21(3):169–179.

Bauer, P. W. (1990). Recent developments in the econometric estimation of frontiers. *Journal of econometrics*, 46(1):39–56.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6):429–444.

Charnes, A., Cooper, W. W., and Rhodes, E. (1981). Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. *Management science*, 27(6):668–697.

Christensen, L. R., Jorgenson, D. W., and Lau, L. J. (1973). Transcendental logarithmic production frontiers. *The review of economics and statistics*, pages 28–45.

Cobb, C. W. and Douglas, P. H. (1928). A theory of production. *The American Economic Review*, 18(1):139–165.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., and Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer Science & Business Media.

Cooper, W. W., Thompson, R. G., and Thrall, R. M. (1996). Introduction: Extensions and new developments in dea. *Annals of Operations Research*, 66(1):1–45.

Cordeiro, J. J., Sarkis, J., Vazquez-Brust, D., Frater, L., and Dijkshoorn, J. (2012). An evaluation of technical efficiency and managerial correlates of solid waste management by welsh smes using parametric and non-parametric techniques. *Journal of the Operational Research Society*, 63(5):653–664.

Diewert, W. E. (1971). An application of the shephard duality theorem: A generalized leontief production function. *The Journal of Political Economy*, pages 481–507.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3):253–290.

Førsund, F. R. (1992). A comparison of parametric and non-parametric efficiency measures: The case of norwegian ferries. In *International Applications of Productivity and*

Efficiency Analysis, pages 21–39. Springer.

Førsund, F. R., Lovell, C. K., and Schmidt, P. (1980). A survey of frontier production functions and of their relationship to efficiency measurement. *Journal of econometrics*, 13(1):5–25.

Gallant, A. R. (1981). On the bias in flexible functional forms and an essentially unbiased form: the fourier flexible form. *Journal of Econometrics*, 15(2):211–245.

Gong, B.-H. and Sickles, R. C. (1992). Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. *Journal of Econometrics*, 51(1):259–284.

Greene, W. H. (1980). Maximum likelihood estimation of econometric frontier functions. *Journal of econometrics*, 13(1):27–56.

Kalirajan, K. P. and Shand, R. T. (1999). Frontier production functions and technical efficiency measures. *Journal of Economic surveys*, 13(2):149–172.

Kumbhakar, S. (2000). C., and cak lovell, stochastic frontier analysis.

Lee, L.-F. (1983). A test for distributional assumptions for the stochastic frontier functions. *Journal of Econometrics*, 22(3):245–267.

McMillan, M. L. and Chan, W. H. (2006). University efficiency: A comparison and consolidation of results from stochastic and non-stochastic methods. *Education economics*, 14(1):1–30.

Meeusen, W. and Van den Broeck, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International economic review*, pages 435–444.

Moutinho, L. and Hutcheson, G. D. (2011). *The SAGE dictionary of quantitative management research*. Sage.

Olesen, O. (1995). Some unsolved problems in data envelopment analysis: A survey. *International journal of production economics*, 39(1):5–36.

Olson, J. A., Schmidt, P., and Waldman, D. M. (1980). A monte carlo study of estimators of stochastic frontier production functions. *Journal of Econometrics*, 13(1):67–82.

Orbanz, P. and Teh, Y. (2010). Encyclopedia of machine learning.

Porcelli, F. (2009). Measurement of technical efficiency. a brief survey on parametric and non-parametric techniques. *University of Warwick*, 11.

Ray, S. C. (2004). *Data envelopment analysis: theory and techniques for economics and operations research*. Cambridge university press.

Richmond, J. (1974). Estimating the efficiency of production. *International economic review*, pages 515–521.

Schaffnit, C., Rosen, D., and Paradi, J. C. (1997). Best practice analysis of bank branches: an application of dea in a large canadian bank. *European Journal of Operational Research*, 98(2):269–289.

Schmidt, P. (1985). Frontier production functions. *Econometric reviews*, 4(2):289–328.

Schmidt, P. et al. (1976). On the statistical estimation of parametric frontier production functions. *The review of economics and statistics*, 58(2):238–39.

Seiford, L. M. (1996). Data envelopment analysis: the evolution of the state of the art (1978–1995). *Journal of productivity Analysis*, 7(2-3):99–137.

Sengupta, J. K. (1987). Data envelopment analysis for efficiency measurement in the stochastic case. *Computers & operations research*, 14(2):117–129.

Simar, L. and Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management science*, 44(1):49–61.

Stevenson, R. E. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of econometrics*, 13(1):57–66.

Thanassoulis, E. (2001). *Introduction to the theory and application of data envelopment analysis*. Springer.

Van den Broeck, J., Koop, G., Osiewalski, J., and Steel, M. F. (1994). Stochastic frontier models: A bayesian perspective. *Journal of Econometrics*, 61(2):273–303.

Winsten, C. (1957). Discussion on mr. farrells paper. *Journal of the Royal Statistical Society*, 120:282–284.